

한국전자책 문서표준(EBKS) 기반 e-Book 의 설계 및 구현

황병연, 심진규
가톨릭대학교 컴퓨터공학과
e-mail: {byhwang, jksim}@catholic.ac.kr

Design and Implementation of e-Book Based on EBKS

Byung Yeon Hwang, Jin Kyu Shim
Dept. of Computer Engineering, Catholic University of Korea

요 약

전자책이란 책의 내용을 디지털 형태로 가공, 저장한 출판물을 의미한다. 인터넷의 급속한 확산에 따른 디지털 콘텐츠 시장의 급성장 속에서 전자책 서비스는 전 세계적으로 많은 관심을 집중시키고 있다. BSP 들의 각기 다른 전자책 문서 포맷의 사용에 따라 전자책 콘텐츠의 교환 및 공유가 어렵게 되자 미국과 일본에 이어 한국에서도 2001 년 한국전자책문서표준(EBKS)이 한국전자책 컨소시엄(EBK)에 의해 제정되었다. 한국전자책문서표준은 OEB(미국전자책표준안), JapaX(일본전자책표준안)와는 달리 고정된 문서의 구조를 정의하여 전자책 콘텐츠의 교환을 보장하고 있으며, XML 을 기본으로 하여 기존의 HTML, PDF 방식의 전자책의 단점을 극복하고 있다. 본 논문에서는 한국전자책 문서표준(EBKS)을 기반으로 웹상에서 유연하게 서비스가 가능한 전자책의 모델을 제시하고, 구현된 전자책을 소개한다.

1. 서론

인터넷으로 다양한 종류의 정보들이 통합되고 사람들은 전세계의 정보를 매우 빠른 속도로 공유할 수 있게 되었다. 사람들이 가장 쉽게 접근할 수 있는 매체인 인터넷을 통한 전자책의 출판은 정보전달 방식에 있어서 큰 변화를 가져올 것이다.

전자책 시장은 시대적인 흐름과 시장의 방대함 속에서 전 세계의 관심을 받고 있다. 현재 수많은 BSP(Book Service Provider)들이 전자책 서비스를 제공하고 있는데, 각자가 다른 포맷을 사용하고 있다. 시장의 변화에 탄력적인 전자책 출판을 위해서는 콘텐츠의 교환과 공유가 필수적인데, 현재 BSP 들은 서로 다른 포맷으로 전자책을 서비스하고 있기 때문에 많은 문제점을 가지고 있다[1].

이러한 전자책 시장의 문제점을 개선하기 위해 미국의 OEBF(Open eBook Forum)와 일본의 JEPA(Japanese Electronic Publishing Association)가 구성되었으며, 1999 년 전자책 문서의 표준안 OEBPS(Open eBook Publication Structure Specification)와 JapaX 0.9 버전을 발표하였다. 국내에서도 한국전자책컨소시엄 EBK(eBook of Korea)[2]가 구성되어 2001 년 5 월 한국전자책문서표준(EBKS:Electronic Book of Korea Standard)을 발표하였다.

EBKS 는 XML 을 기본 포맷으로 하여 전자책의 콘텐츠와 스타일을 분리하고, 콘텐츠와 스타일 정보 모두를 교환의 대상으로 하고 있다. 문서의 구조를 XML DTD 를 이용해 명확히 기술하여 고정함으로써 BSP 간의 교환과 공유가 원만히 이루어지도록 하고 있으며, 전자책 콘텐츠의 출력을 위한 논리 구조와 스타일 정보는 XSL-FO 의 사용을 권고하고 있다. 그러나 XSL-FO 는 아직 개발 단계에 있기 때문에 현실적에서 전자책 출판을 위해 적용하기에는 문제점을 가지고 있다.

본 연구에서는 스타일 언어로 XSL-FO 를 적용하기 어려운 현실을 고려하여 기존의 CSS 를 사용한 전자책 모델을 소개한다. 표준안이 문서의 구조를 XML DTD 로 명확히 하고 있다는 점을 이용하여 책의 목차 보여주기, 이전, 다음 페이지 보여주기, 원하는 페이지 보여주기, 키워드의 검색 기능 등을 DOM 을 이용한 XML 문서의 HTML 변환 모델로 구현한다. 본 연구가 제시하는 모델로 전자책 출판을 하게 될 경우, XSL-FO 를 스타일 언어로 사용할 수 있는 시점이 될 때 문서의 스타일 출력 부분만을 수정하고, 콘텐츠 문서는 재사용 할 수 있으므로 비용의 절감효과를 기대할 수 있을 것이다

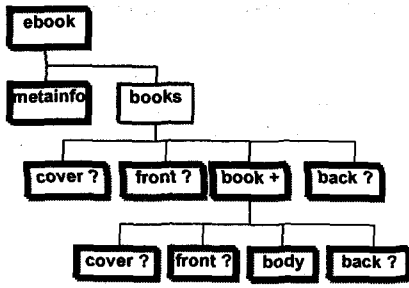
2. 한국전자책 표준안(EBKS)

2.1 EBKS 가 채택하고 있는 표준안

EBKS 는 메타데이터로 더블린코어[3]의 15 개 기본 엘리먼트를 제공한다. 또한 전자책 업계에서의 다양한 응용을 고려하여 확장 가능한 x-metadata 를 지원한다. 한글의 인코딩은 OEB 와 JapaX 와의 호환을 고려하여 Unicode 를 기본으로 하고 있고, UTF-8, UTF-16 도 지원하도록 한다. 언어표기는 DC 의 'Language'와 RTF1766 을, 언어 이름은 ISO 630 표준을 따른다. 또한 x-character 셋을 제공하여 EBKS entity 에 정의되지 않은 특수문자, 한자, 고어 등의 표기를 지원한다.

2.2 EBKS DTD 의 구조

EBKS DTD[1]는 [그림 1]과 같은 구조를 가지고 있다.

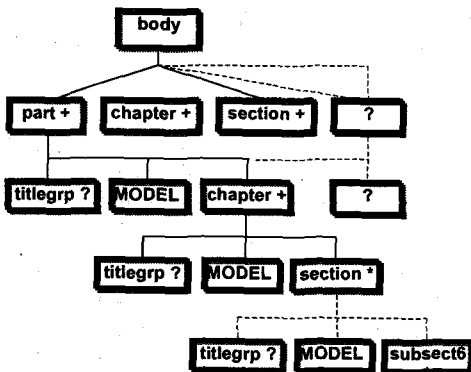


[그림 1] EBKS DTD 의 구조

루트 엘리먼트인 eBook 은 메타데이터 표현을 위한 metainfo 엘리먼트와 books 엘리먼트로 구성된다.

books 엘리먼트는 이미지, 출판사에 대한 상세 정보들을 표현하는 cover 엘리먼트, 머리말, 감사의 글, 서문, 목차 등을 표현하는 front 엘리먼트, 책의 내용을 표현하는 book 엘리먼트 등으로 구성된다. books 엘리먼트는 하나 이상의 book 엘리먼트를 포함할 수 있는데, 이는 전집류, 모음집 형태의 서비스를 지원하기 위함이다.

book 엘리먼트는 cover, front, body, back 엘리먼트들을 포함할 수 있다. body 엘리먼트는 [그림 2]와 같은 구조를 가지고 있다.



[그림 2] EBKS DTD- body element 의 구조

body 엘리먼트는 책의 실질적인 내용을 기술하는 부분으

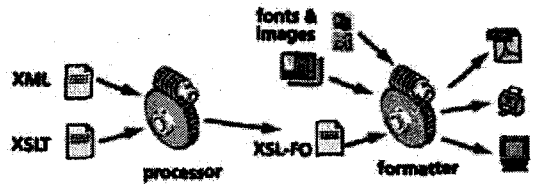
로 part, chapter, section 엘리먼트를 포함한다. 대부분의 본문 구조는 part 엘리먼트를 이용하여 상위 레벨의 구조를 정의하므로 part 엘리먼트는 매우 중요하게 사용될 수 있다. chapter 엘리먼트는 일반적으로 어떤 한 주제가 일단락이 되지만 전후와의 연관성을 갖고 있는 책의 한 부분을 표현하는데 사용된다. section 엘리먼트는 총 6 레벨의 subject 엘리먼트로 구성되며, 이는 대부분의 경우 6 레벨의 subsection 구조로 처리가 가능하기 때문이다.

2.3 EBKS 의 스타일

EBKS 에서는 전자책 콘텐츠 저작자들의 의도를 다양한 환경에서 일관성 있는 출력 결과형태로 나타내기 위해 스타일 정보를 교환의 대상으로 삼는다. 이를 위하여 XSL-FO 를 권고하고 있다. XSL-FO 는 기존의 CSS 및 CSS2 에 비하여 그 기능이 다양하고, 전자 출판물에 대한 출력을 고려한 사양이기 때문이다.

2.4 XSL-FO

인터넷문서 표준인 XML(Extensible Markup Language)[4] 이 구조적인 내용만을 갖고 있기 때문에 문서를 보여주기 위한 표현 정보를 포함하는 스타일 시트(style sheets)가 필요하다. 이를 위해 W3C(World Wide Web Consortium)에서는 XML 문서의 구조적인 접근을 위한 XPath(XML Path Language)와 문서의 내용구조를 변환하기 위한 XSLT(XSL Transformations), 그리고 포매팅 정보를 위한 XSL-FO(XSL Formatting objects)로 구성되는 XSL(Extensible Stylesheet Language)을 제안하였다[5]. XSL 의 XML 처리과정은 [그림 3]과 같다.



[그림 3] XSL 처리과정 : XSLT 변환, Formatting

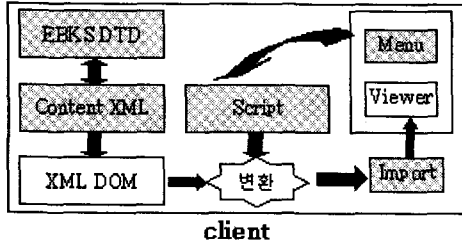
XSL-FO 는 단순한 웹 문서뿐만 아니라 전자책의 출판과 같은 고품위 문서에 대한 출력스타일의 정의도 가능하다. 그러나 XSL-FO 는 출현한지 1년 반도 안된 새로운 언어이기 때문에 웹 환경에 아직 적용이 되지 않은 상황이다. CSS 의 웹 환경 적응기간이 4 년이었고, 새로 등장하는 언어의 웹 적응 속도가 점차 늦어지고 있는 현실을 감안할 때, XSL-FO 를 지원하는 웹 브라우저가 시판되려면 3-5 년의 시간이 필요할 것으로 추정된다. 현재 XSL-FO 를 지원하는 웹 브라우저가 아직 시판되지 않고 있으며 일부 제한된 기능을 하는 브라우저로 Apache 의 FOP, TEI 의 PassiveTeX, AntennaHouse 의 XSL Formatter, RenderX 의 XEP, Arbortext 의 E3, IBM 의 Infoprint XML Extender for z/OS, Adobe Systems 의 Adobe Document Server 5.0 등이 나와있다[6,7].

3. 전자책의 설계 및 구현

3.1 전자책의 설계

본 연구에서 구현한 전자책은 EBKS DTD 표준안과 연동되

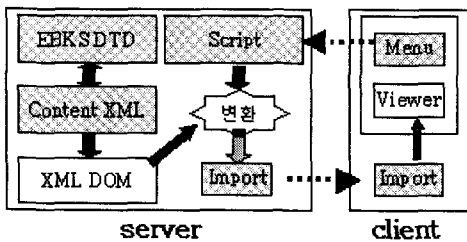
는 콘텐츠 XML 데이터 문서, 사용자에게 제공되는 메인 HTML, XML DOM 을 조작하기 위한 스크립트, 변환된 HTML 문서가 출력될 import HTML로 구성되어 있다.



[그림 4] 전자책 모델

사용자에게 제공되는 메인 HTML 문서는 입력부분과 변환 부분, 출력부분으로 나누어 볼 수 있다. 입력부인 Menu 부분에서는 사용자가 원하는 책의 정보를 입력 받고, 외부의 스크립트에 존재 하는 함수를 호출한다. 변환부에서 호출된 함수는 입력 받은 정보를 이용하여 메모리상에 존재하는 전자책 XML 데이터 문서의 Document Object Model 에 접근하여 사용자가 원하는 부분을 추출하여 HTML 문서로 변환한다. 출력부에서는 변환부에서 가공된 HTML 문서를 메인 HTML 문서의 Viewer 부분에 출력한다.

인터넷에서 서비스되고 있는 전자책들은 온라인 상태에서 서비스를 제공하는 방식과, BSP 가 개발한 자체 뷰어와 데이터 문서를 다운받아 오프라인 상태에서 사용하는 방식으로 나누어 볼 수 있다. 본 연구에서 제안한 전자책의 모델은 [그림 4]와 같이 입력, 변환, 출력부를 클라이언트로 모두 전송하여 오프라인 상태에서 사용할 수 있으며, [그림 5]와 같이 변환부를 서버측에 두고 변환된 HTML 문서만을 클라이언트로 전송하는 형태의 온라인 모델로도 사용할 수 있다.



[그림 5] 온라인 전자책 서비스 모델

기존의 온라인 형태로 제공되는 전자책은 HTML 포맷을 사용하는데, HTML 의 형태로 전자책을 서비스 하게 될 경우 문서의 데이터 부분과 스타일 부분이 분리되어있지 않았기 때문에 데이터 문서의 재사용이 불가능했다. 따라서 책의 내용을 있는 그대로 보여주기만 할 뿐 사용자가 원하는 형태로 데이터를 가공하여 보여주는 것이 어려웠다. 본 연구에서 구현한 전자책은 전자책 XML 데이터문서를 DOM 변환을 이용하여 다양한 방법으로 출력할 수 있도록 하였으며, 콘텐츠 XML 데이터 문서를 EBKS 에 적합한 XML 문서로 구축함으로써 BSP 간에 데이터 문서의 교환이 용이하도록 하였다.

3.2 전자책의 구현

본 연구의 전자책 구현에서는 데이터 문서의 DTD 로 EBKS DTD version 1.0 Draft 를, 데이터 문서로는 대한 성서공회회의 '공동번역 성서' 중 신약성서 부분을, 데이터 문서가 DTD 에 유효한지의 여부는 Microsoft 사의 'XML Validation Tool'을 사용하였다. 웹 브라우저는 'Microsoft Internet Explorer 6.0'을, XML Parser 로는 'MSXML 3.0'을, 스크립트 언어는 'java script'를, 스타일 언어는 'CSS'를 사용하였다.

데이터 문서로 사용한 신약성서는 '책 제목', '복음서 제목', '장 이름', '소제목', '절의 내용'으로 구성된다. EBKS DTD 에 적합한 구조로 작성한 XML 문서의 내용은 [그림 6]과 같다. 작성한 데이터 문서와 EBKS DTD 의 유효성 여부는 'XML Validation Tool'을 사용하여 검증하였다.

```

-Ebook/ books/ book/ body/
  part/ titlegrp/ title/ 복음서 제목
-Ebook/ books/ book/ body/
  part/ chapter/ section/
  titlegrp/ title/ 장 제목
-Ebook/ books/ book/ body/
  part/ chapter/ section/ subsect1/
  titlegrp/ title/ 소제목
-Ebook/ books/ book/ body/
  part/ chapter/ section/ subsect1/
  p/ 절의 내용
    
```

[그림 6] 전자책 XML 데이터

입력부분인 Menu 부분에서는 목차 보기, 선택한 복음서, 선택한 장 보기, 다음장 보기, 이전장 보기, 지정한 검색 범위 내에서 키워드 검색하기, 책 전체에서 키워드 검색하기 기능을 구현하여 전자책 사용자에게 다양한 형태로 책의 정보를 제공할 수 있도록 하였다.

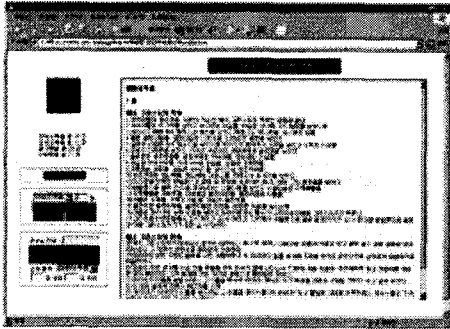
데이터 XML 문서의 각 엘리먼트들은 노드의 형태로 메모리상에 트리 구조로 존재하게 된다. 데이터 문서는 EBKS DTD 가 정의한 정형화된 구조에 텍스트 노드만 추가되어 표현되므로 그 구조가 규칙적이다. 입력부에 의해 호출된 외부 스크립트의 해당 함수는 규칙적인 구조의 콘텐츠 XML 문서의 DOM 구조를 파악하고 필요로 하는 텍스트 노드를 취하여 사용자가 원하는 형태의 HTML 문서로 변환한다. DOM 에 존재하는 각 노드들의 접근 방법은 [그림 7]과 같다. aa, bb, cc, dd 는 복음서 제목, 장 제목, 각 장의 소제목, 절의 내용 각각에 접근할 수 있도록 하는 인자로 사용된다.

```

-Ebook/ books :
var rr = myXML.documentElement.childNodes(1).
  childNodes(0).childNodes(2)
- 복음서 제목 :
  rr.childNodes(aa).childNodes(0).childNodes(0).text
- 장 제목 :
  rr.childNodes(aa).childNodes(1).childNodes(bb).
  childNodes(0).childNodes(0).text
- 각 장의 소제목 :
  rr.childNodes(aa).childNodes(1).childNodes(bb).
  childNodes(cc).childNodes(0).childNodes(0).text
- 절 내용 :
  rr.childNodes(aa).childNodes(1).childNodes(bb).
  childNodes(cc).childNodes(dd).text
    
```

[그림 7] DOM 의 접근

마태오 복음 1 장을 선택하고 '선택한 장 보기'를 클릭하면, 스크립트의 해당 함수는 메모리상에 존재하는 DOM 에서 마태오 복음 1 장 부분만을 취하여 HTML 문서로 변환하여 임포트될 HTML 문서에 삽입한다. 변환된 문서가 삽입된 HTML 문서는 변환된 문서를 메인 페이지의 뷰어 화면에 임포트 한다. [그림 8]은 결과 출력 화면이다.



[그림 8] 보기 출력화면

검색키워드에 '죽보'를, 검색 범위에 '마태오', '1~2 장'을 입력하고 '검색범위에서 찾기'를 클릭하면 스크립트의 해당 함수는 DOM 에서 마태오 복음의 1 장~2 장 부분을 취하고 그 안에서 해당 키워드가 있는지를 검색하여 그 결과를 HTML 로 변환한다. [그림 9]는 검색 입력 폼을, [그림 10]은 뷰어 화면에 출력된 변환된 문서이다.

검색할 단어 :

검색 범위 :

장 부터 장 까지

[그림 9] 검색 입력 폼

지정된 범위에서 키워드 검색결과 총 2개가 검색되었습니다.

- 1. [마태오 1 장-1절] 소자목) 예수 그리스도의 죽보
- 2. [마태오 1 장-1절] 마태오의 후손이요, 다윗의 자손인 예수 그리스도의 죽보는 다음과 같다.

[그림 10] 검색 결과

4. 결론

현재 전자책은 컴퓨터 기술의 발전 및 인터넷의 급속한 확산에 따른 디지털 콘텐츠 시장의 급성장, 그리고 전자책이 갖는 장점이 복합적으로 결합되어 전 세계적으로 많은 관심을 집중시키고 있다. 이러한 시점에서 한국전자책컨소시엄(EBK)은 한국전자책표준(EBKS)을 발표하여 BSP 간의 콘텐츠의 교환과 공유를 시도하고 있다. 그러나 표준안이 권고하고 있는 스타일 언어 XSL-FO 는 개발중인 언어이기 때문에 BSP 들이 이를 활용하지 못하고 있다. XSL-FO 를 사용할 수 있는 시점이 올때까지 BSP 들이 각자의 포맷으

로 전자책 콘텐츠를 개발하여 사용할 경우 나중에 데이터의 교환과 공유를 위해 EBKS 기반의 데이터로 변환해야 할 것이다. 본 논문에서는 EBKS 기반의 전자책 XML 데이터문서와 CSS 를 사용하여 현 시점에서 서비스가 가능한 EBKS 기반의 전자책을 소개하였다. 또한 DOM 변환을 이용하여 전자책 사용자들의 다양한 욕구를 충족시킬 수 있도록 하였다. 본 연구가 제시하는 모델로 전자책 출판을 하게 될 경우, XSL-FO 를 스타일 언어로 사용할 수 있는 시점이 될 때 문서의 스타일 출력 부분만을 수정하고, 콘텐츠 문서는 재사용 할 수 있으므로 비용의 절감효과를 기대할 수 있을 것이다.

참고문헌

- [1] 손원성, 고승규, 이경호, 김성혁, 임순범, 최윤철, " XML 에 기반한 한국 전자책 문서 표준", 한국정보처리학회지, Vol.8, No.3, 2001.
- [2] 한국전자책 컨소시엄, <http://www.ebk.or.kr>
- [3] Dublin Core Metadata Element Set, Version 1.1: Reference Description, <http://purl.org/dc/elements/1.1/>
- [4] W3C Consortium, " Extensible Markup Language(XML) 1.0", <http://www.w3c.org/TR/1998/RECxml-19980210>, 1998.
- [5] Stephan Deach, " What Is XSL-FO and When Should I Use It?", The Seybold Report Analyzing Publishing Technologies, Vol.2, No.17, 2002.
- [6] 손원성, 고승규, 이경호, 김성혁, 임순범, 최윤철, " 한국전자책 문서표준(EBKS)의 개발 및 활용방안", 한국정보과학회 2001년 춘계학술대회, 2001.
- [7] Antenna House, Antenna House XSL Formatter, <http://www.antenna.co.jp/>, 2001.